

# Multi-Shift de Bruijn Sequence

Zhi Xu\*

The University of Western Ontario,  
Department of Computer Science,  
London, Ontario, Canada N6A 5B7  
zhi\_xu@csd.uwo.ca

## Abstract

A (non-circular) de Bruijn sequence  $w$  of order  $n$  is a word such that every word of length  $n$  appears exactly once in  $w$  as a factor. In this paper, we generalize the concept to a multi-shift setting: a  $m$ -shift de Bruijn sequence of order  $n$  is a word such that every word of length  $n$  appears exactly once in  $w$  as a factor that starts at an index  $im + 1$  for some integer  $i \geq 0$ . We show the number of the  $m$ -shift de Bruijn sequences of order  $n$  is  $a^n!a^{(m-n)(a^n-1)}$  for  $1 \leq n \leq m$  and is  $(a^m!)^{a^{n-m}}$  for  $1 \leq m \leq n$ , where  $a$  is the size of the alphabet. We provide two algorithms for generating a multi-shift de Bruijn sequence. The multi-shift de Bruijn sequence is important in solving the Frobenius problem in a free monoid.

## 1 Introduction

If a word  $w$  can be written as  $w = xyz$ , then words  $x$ ,  $y$ , and  $z$  are called the prefix, factor, and suffix of  $w$ , respectively. A word  $w$  over  $\Sigma$  is called a de Bruijn sequence of order  $n$ , if each word in  $\Sigma^n$  appears exactly once in  $w$  as a factor. For example, 00110 is a binary de Bruijn sequence of order 2 since each binary word of length two appears in it exactly once as a factor:  $00110 = (00)110 = 0(01)10 = 00(11)0 = 001(10)$ . The de Bruijn sequence can be understood by the following game. Suppose there are infinite supplies of balls, each of which is labeled by a letter in  $\Sigma$ , and there is a glass pipe that can hold balls in a vertical line. On the top of that pipe is an opening, through which one can drop balls into that pipe, and on the bottom is a trap-door, which can support the weight of at most  $n$  balls. When there are more than  $n$  balls in the pipe, the trap-door opens and those balls at the bottom drop off until only  $n$  balls remain. If we put balls as numbered as in a de Bruijn sequence on the alphabet  $\Sigma$  of order  $n$ , then every  $n$  ball sequence will appear exactly once in the pipe. It is easy to see that a de-Bruijn sequence of order  $n$ , if exists, is of length  $|\Sigma|^n + n - 1$  and its suffix of length  $n - 1$  is identical to its prefix of length  $n - 1$ . So, sometimes a de-Bruijn sequence is written in a circular form by omitting the last  $n - 1$  letters, which can be viewed as the equivalence class of words under the conjugate relation.

The de Bruijn sequence is also called the de Bruijn-Good sequence, named after de Bruijn [2] and Good [7] who independently studied the existence of such words over binary alphabet; the former also provided a formula  $2^{2^n}$  for the total number of those words of order  $n$ . The study of the de Bruijn sequence, however, dates back at least to 1894, when Flye Sainte-Marie [3] studied the

---

\*Part of the work was done during the author's stay at University of Waterloo

words and provided the same formula  $2^{2^n}$ . For an arbitrary alphabet  $\Sigma$ , van Aardenne-Ehrenfest and de Bruijn [1] provided the formula  $(|\Sigma|!)^{|\Sigma|^n}$  for the total number of de Bruijn sequences of order  $n$ . Besides the total number of de Bruijn sequences, another interesting topic is how to generate a de Bruijn sequence (arbitrary one, lexicographically least one, lexicographically largest one). For generating de Bruijn sequences, see the surveys [4, 12]. The de Bruijn sequence is some times called the full cycle [4], and has connections to the following concepts: feedback shift registers [6], normal words [7], generating random binary sequences [10], primitive polynomials over a Galois field [13], Lyndon words and necklaces [5], Euler tours and spanning trees [1].

In this paper, we consider a generalization of the de Bruijn sequence. To understand the concept, let us return to the glass pipe game presented at the beginning. Now the trap-door can support more weight. When there are  $n + m$  or more balls in the pipe, the trap-door opens and the balls drop off until there are only  $n$  balls in the pipe. Is there an arrangement of putting the balls such that every  $n$  ball sequence appears exactly once in the pipe? The answer is “Yes” for arbitrary positive integers  $m, n$ . The solution represents a multi-shift de Bruijn sequence. We will discuss the existence of the multi-shift de Bruijn sequence, the total number of multi-shift de Bruijn sequences, generating a multi-shift de Bruijn sequence, and the application of the multi-shift de Bruijn sequence in the Frobenius problem in a free monoid.

## 2 Multi-Shift Generalization of the de Bruijn Sequence

Let  $\Sigma \subseteq \{0, 1, \dots\}$  be the *alphabet* and let  $w = a_1 a_2 \dots a_n$  be a word over  $\Sigma$ . The *length* of  $w$  is denoted by  $|w| = n$  and the *factor*  $a_i \dots a_j$  of  $w$  is denoted by  $w[i..j]$ . If  $u = w[im + 1..im + n]$  for some non-negative integer  $i$ , we say factor  $u$  appears in  $w$  at a *modulo  $m$  position*. The set of all words of length  $n$  is denoted by  $\Sigma^n$  and the set of all finite words is denoted by  $\Sigma^* = \{\epsilon\} \cup \Sigma \cup \Sigma^2 \dots$ , where  $\epsilon$  is the *empty word*. The concatenation of two words  $u, v$  is denoted by  $u \cdot v$ , or simply  $uv$ .

A word  $w$  over  $\Sigma$  is called a *multi-shift de Bruijn sequence* of shift  $m$  and order  $n$ , if each word in  $\Sigma^n$  appears exactly once in  $w$  as a factor at a modulo  $m$  position. For example, one of the 2-shift de Bruijn sequence of order 3 is 00010011100110110, which can be verified as follows:

$$\begin{aligned} 00010011100110110 &= (000)10011100110110 = 00(010)011100110110 \\ &= 0001(001)1100110110 = 000100(111)00110110 = 00010011(100)110110 \\ &= 0001001110(011)0110 = 000100111001(101)10 = 00010011100110(110). \end{aligned}$$

The multi-shift de Bruijn sequence generalizes the de Bruijn sequence in the sense de Bruijn sequences are exactly 1-shift de Bruijn sequences of the same order. It is easy to see that the length of each  $m$ -shift de Bruijn sequence of order  $n$ , if exists, is equal to  $m|\Sigma|^n + (n - m)$ . By the definition of multi-shift de Bruijn sequence, the following proposition holds.

**Proposition 1.** *Let  $w$  be one  $m$ -shift de Bruijn sequence  $w$  of order  $n$ ,  $n > m$ . Then the suffix of length  $n - m$  of  $w$  is identical to the prefix of length  $n - m$  of  $w$ .*

*Proof.* Let  $w$  be one  $m$ -shift de Bruijn sequence  $w$  of order  $n$  over  $\Sigma$  and let  $a = |\Sigma|$ . Write  $n = km + r$  such that  $0 < r \leq m$ . If  $k = 1$ , then we compare the set of all factors  $w[(i + 1)m + 1..(i + 1)m + r]$  and the set of all factors  $w[im + 1..im + r]$  for  $0 \leq i \leq a^n - 1$ . The former covers factors  $u[m + 1..m + r]$  and the latter covers factors  $u[1..r]$  for every  $u \in \Sigma^n$ . Since the two are identical, we have  $w[a^n m + 1..a^n m + r] = w[1..r]$ . Now we assume  $k \geq 2$ . Consider the set of all factors  $w[(i + j + 1)m + 1..(i + j + 2)m]$  and the set of all factors  $w[(i + j)m + 1..(i + j + 1)m]$  for

$0 \leq i \leq a^n - 1$  and  $0 \leq j < k$ . By the same argument, we have  $w[(a^n + j)m + 1 .. (a^n + j + 1)m] = w[jm + 1 .. (j + 1)m]$  for  $0 \leq j < k$ . Finally, comparing the set of all  $w[(i + k)m + 1 .. (i + k + 1)m + r]$  and the set of all  $w[(i + k - 1)m + 1 .. (i + k - 1)m + r]$  for  $0 \leq i \leq a^n - 1$ , we have the equality  $w[(a^n + k - 1)m + 1 .. (a^n + k - 1)m + r] = w[(k - 1)m + 1 .. (k - 1)m + r]$ . Therefore, we have the equality  $w[ma^n + 1 .. ma^n + n - m] = w[a^n m + 1 .. (a^n + k - 1)m + r] = w[1 .. (k - 1)m + r] = w[1 .. n - m]$ .  $\square$

From Proposition 1, we know that when  $n > m$ , every multi-shift de Bruijn sequence can be written as a circular word and the discussion on multi-shift de Bruijn sequences of the two different forms are equivalent. In this paper, we discuss the multi-shift de Bruijn sequence in the form of ordinary words.

A (non-strict) directed graph, or digraph for short, is a triple  $G = (V, A, \psi)$  consisting of a set  $V$  of vertices, a set  $A$  of arcs, and an incidence function  $\psi : A \rightarrow V \times V$ . Here we do not take the convention  $A \subseteq V \times V$ , since we allow a digraph contains self-loops and multiple arcs regarding the same pair of vertices. When  $\psi(a) = (u, v)$ , we say the arc  $a$  joins  $u$  to  $v$ , where vertex  $u = \text{tail}(a)$  and vertex  $v = \text{head}(a)$  are called *tail* and *head*, respectively. The indegree  $\delta^-(v)$  (outdegree  $\delta^+(v)$ , respectively) of a vertex  $v$  is the number of arcs with  $v$  being the head (the tail, respectively). A walk in  $G$  is a sequence  $a_1, a_2, \dots, a_k$  such that  $\text{head}(a_i) = \text{tail}(a_{i+1})$  for each  $1 \leq i < k$ . The walk is closed, if  $\text{head}(a_k) = \text{tail}(a_0)$ . Two closed walks are regarded as identical if one is the circular shift of the other. An Euler tour is a closed walk that traverses each arc exactly once. A Hamilton cycle is a closed walk that traverses each vertex exactly once. An (spanning) arborescence is a digraph with a particular vertex, called the root, such that it contains every vertices of  $G$ , its number of arcs is exactly one less than the number of vertices, and there is exactly one walk from the root to any other vertex. We denote the total number of Euler tours, Hamilton cycles, and arborescence of  $G$  by  $|G|_E$ ,  $|G|_H$ , and  $|G|_A$ , respectively.

An (undirected) graph is defined as a digraph such that for any pair of vertices  $v_1, v_2$ , there is an arc  $a$ ,  $\psi(a) = (v_1, v_2)$ , if and only if there is a corresponding arc  $a'$ ,  $\psi(a') = (v_2, v_1)$ . In this case, we write  $\delta^-(v) = \delta^+(v) = \delta(v)$  and a spanning arborescence is just a spanning tree.

The arc-graph  $G^*$  of  $G = (V, A, \psi)$  is defined as  $(A, C, \varphi)$  such that for every pair of arcs  $a_1, a_2 \in A$ ,  $\text{head}(a_1) = \text{tail}(a_2)$ , there is an arc  $c \in C$ ,  $\varphi(c) = (a_1, a_2)$  and those arcs are the only arcs in  $C$ . Euler tours exist in a graph  $G$  if and only if Hamilton cycles exist in the arc-graph  $G^*$ .

We define the word graph  $G(m, n)$  by  $(\Sigma^n, \Sigma^{n+m}, \psi)$ , where  $\psi(w) = (u, v)$  for  $u = w[1 .. n]$ ,  $v = w[m + 1 .. m + n]$ . Then by definition, the following lemmas are straightforward.

**Lemma 2.** *The digraph  $G(m, n)^*$  is the digraph  $G(m, n + m)$ .*

*Proof.* By definition,  $G(m, n) = (\Sigma^n, \Sigma^{n+m}, \psi')$ ,  $G(m, n + m) = (\Sigma^{n+m}, \Sigma^{n+2m}, \psi'')$ , where  $\text{tail}'(w) = w[1 .. n]$ ,  $\text{head}'(w) = w[m + 1 .. m + n]$ , and  $\text{tail}''(w) = w[1 .. m + n]$ ,  $\text{head}''(w) = w[m + 1 .. 2m + n]$ . So for every pair of arcs  $a_1, a_2 \in \Sigma^{n+m}$  of  $G(m, n)$  with  $\text{head}'(a_1) = \text{tail}'(a_2)$ , there is an arc  $a_1 \cdot a_2[n + 1 .. n + m] \in \Sigma^{n+2m}$  of  $G(m, n + m)$ ; and for every arc  $w \in \Sigma^{n+2m}$  of  $G(m, n + m)$ ,  $\text{head}'(\text{tail}''(w)) = w[m + 1 .. m + n] = \text{tail}'(\text{head}''(w))$ . Hence, by definition,  $G(m, n + m)$  is the arc-graph of  $G(m, n)$ .  $\square$

**Lemma 3.** *Suppose  $m \leq n$ . (1) There is a  $|\Sigma|^n$ -to-1 mapping from the set of  $m$ -shift de Bruijn sequences of order  $n$  onto the set of Hamilton cycles in  $G(m, n)$ . (2) There is a  $|\Sigma|^n$ -to-1 mapping from the set of  $m$ -shift de Bruijn sequences of order  $n$  onto the set of Euler tours in  $G(m, n - m)$ .*

*Proof.* Let  $l = |\Sigma|^n$ . (1) Notice that any Hamilton cycle  $a_1, a_2, \dots, a_l$  together with a starting arc  $a_1$  uniquely determines one  $m$ -shift de Bruijn sequences of order  $n$  specified by

$$a_1[1..n]a_1[n+1..n+m]a_2[n+1..n+m]\cdots a_{l-1}[n+1..n+m],$$

and vice versa. So the  $l$ -to-1 mapping exists. (2) Applying Lemma 2, this part follows from (1).  $\square$

**Theorem 4.** *For any alphabet  $\Sigma$ , positive integers  $m, n$ , the  $m$ -shift de Bruijn sequences of order  $n$  over  $\Sigma$  exist.*

*Proof.* First we assume  $m \geq n$ . Let  $u_1, u_2, \dots, u_l$  be any permutation of the words in  $\Sigma^n$  for  $l = |\Sigma|^n$ . Then the word  $u_1 0^{m-n} u_2 0^{m-n} \dots 0^{m-n} u_l$  is one  $m$ -shift de Bruijn sequence of order  $n$  over  $\Sigma$ .

Now we assume  $m < n$  and prove there exists an Euler tour in  $G(m, n-m)$ . Then by Lemma 3, the existence of  $m$ -shift de Bruijn sequences of order  $n$  over  $\Sigma$  is ensured. To show the existence of an Euler tour, we only need to verify that  $G(m, n-m)$  is connected and that  $\delta^-(v) = \delta^+(v)$  for every vertex  $v$ , both of which are straightforward: for every vertex  $v$  in  $G(m, n-m)$ ,  $v$  is connected to the vertex  $0^{n-m}$  in both directions and  $\delta^-(v) = \delta^+(v) = |\Sigma|^m$ .  $\square$

### 3 Counting the Number of Multi-Shift de Bruijn Sequences

Since  $m$ -shift de Bruijn sequence of order  $n$  exists, in this section we discuss the total number of different  $m$ -shift de Bruijn sequence of order  $n$ , and we denote the number by  $\#(m, n)$ . First, we study the degenerated case.

**Lemma 5.** *For  $1 \leq n \leq m$ ,  $\#(m, n) = a^n! a^{(m-n)(a^n-1)}$ , where  $a = |\Sigma|$ .*

*Proof.* Let  $a = |\Sigma|$ . By the definition of the multi-shift de Bruijn sequence, in the case  $1 \leq n \leq m$ ,  $m$ -shift de Bruijn sequences of order  $n$  are exactly those of the form  $u_1 \Sigma^{m-n} u_2 \Sigma^{m-n} \dots \Sigma^{m-n} u_l$ , where  $l = a^n$  and  $u_1, u_2, \dots, u_l$  is a permutation of all words in  $\Sigma^n$ . Therefore, the total number of such words is  $a^n! a^{(m-n)(a^n-1)}$ .  $\square$

To study the case  $1 \leq m \leq n$ , we need a theorem by van Aardenne-Ehrenfest and de Bruijn [1], which describes the relation between the number of Euler tours in a particular type of digraph and the number of Euler tours in its arc-graph.

**Theorem 6** (van Aardenne-Ehrenfest and de Bruijn [1]). *Let  $G = (V, A, \psi)$  be a digraph such that  $a = \delta^-(v) = \delta^+(v)$  for every  $v \in V$ . Then  $|G^*|_E = a^{-1}(a!)^{|V|}(a-1)|G|_E$ .*

The digraph  $G(m, n)$  satisfies the conditions in Theorem 6 with  $a = |\Sigma|^m$ . So, by the relation between the multi-shift de Bruijn sequences and the Euler tours in the word graph  $G(m, n)$ , we have the following recursive expression on  $\#(m, n)$ .

**Lemma 7.** *For  $m \geq 1, n \geq 2m$ ,  $\#(m, n) = (a^m!)^{a^{n-m}-a^r} \#(m, m+r)$ , where  $a = |\Sigma|$ ,  $r = n \bmod m$ .*

*Proof.* Let  $a = |\Sigma|$ ,  $r = n \bmod m$ . By Lemma 3,

$$\begin{aligned}
\#(m, n) &= a^n |G(m, n-m)|_E \\
&= a^{n-m} (a^m!)^{a^{n-2m}(a^m-1)} |G(m, n-2m)|_E \\
&= (a^m!)^{a^{n-2m}(a^m-1)} \#(m, n-m) \\
&= (a^m!)^{a^{n-2m}(a^m-1)} (a^m!)^{a^{n-3m}(a^m-1)} \#(m, n-2m) \\
&= \dots \\
&= (a^m!)^{a^{n-2m}(a^m-1)} (a^m!)^{a^{n-3m}(a^m-1)} \dots (a^m!)^{a^r(a^m-1)} \#(m, m+r) \\
&= (a^m!)^{a^{n-m}-a^r} \#(m, m+r). \quad \square
\end{aligned}$$

To finish the last step of obtaining  $\#(m, n)$  for  $1 \leq m \leq n$ , we again need two theorems, which are often used in the literature to count the number of Euler tours in various types of digraphs.

**Theorem 8** (BEST theorem [14, 1]). *In a digraph  $G = (V, A, \psi)$ ,  $|G|_E = \prod_{v \in V} (\delta^+(v) - 1)! |G|_A$ .*

**Theorem 9** (Kirchhoff's matrix tree theorem [9]). *In a graph  $G = (V, A, \psi)$ , the number of spanning trees is equal to any cofactor of the Laplacian matrix of  $G$ , which is the diagonal matrix of degrees minus the adjacency matrix.*

**Lemma 10.** *For  $1 \leq m \leq n \leq 2m$ ,  $\#(m, n) = (a^m!)^{a^{n-m}}$ , where  $a = |\Sigma|$ .*

*Proof.* Let  $r = n-m$  and  $a = |\Sigma|$ . Then  $0 \leq r \leq m$ . By definition,  $G = G(m, n-m) = (\Sigma^r, \Sigma^m, \psi)$ . So from any vertex to any vertex, there are  $a^{m-r}$ -many arcs in  $G$ . We convert  $G$  into a undirected graph  $G'$  by omitting all self-loops; there are  $a^{m-r}$ -many of them for each vertex. Since for every pair of vertices  $v_1, v_2$  there are  $a^{m-r}$ -many arcs joins  $v_1$  to  $v_2$  and correspondingly there are  $a^{m-r}$ -many arcs joins  $v_2$  to  $v_1$ , the graph  $G'$  is indeed an undirected graph by our definition. Each vertex in  $G'$  is of degree  $a^m - a^{m-r}$ . Then the Laplacian matrix of  $G'$  is

$$L = \begin{pmatrix} a^m - a^{m-r} & -a^{m-r} & \dots & -a^{m-r} \\ -a^{m-r} & a^m - a^{m-r} & \dots & -a^{m-r} \\ \vdots & \vdots & \ddots & \vdots \\ -a^{m-r} & -a^{m-r} & \dots & a^m - a^{m-r} \end{pmatrix}.$$

By Theorem 9, the number of arborescence  $|G|_A = |G'|_A$  is equal to the cofactor of  $L$ , which is  $(a^m)^{a^r-2} a^{m-r} = (a^m)^{a^r} / a^n$ . Then by Theorem 8, the number of Euler tours in digraph  $G$  is  $|G|_E = ((a^m - 1)!)^{a^r} |G|_A = ((a^m - 1)!)^{a^r} (a^m)^{a^r} / a^n = (a^m!)^{a^r} / a^n$ . Finally, by Lemma 3, the number of  $m$ -shift de Bruijn sequence of order  $n$  is  $\#(m, n) = a^n |G|_E = (a^m!)^{a^r}$ .  $\square$

**Theorem 11.** *For  $1 \leq n \leq m$ ,  $\#(m, n) = a^n! a^{(m-n)(a^n-1)}$ , and for  $1 \leq m \leq n$ ,  $\#(m, n) = (a^m!)^{a^{n-m}}$ , where  $a = |\Sigma|$ .*

*Proof.* For  $1 \leq n \leq m$ , the equality  $\#(m, n) = a^n! a^{(m-n)(a^n-1)}$  is shown in Lemma 5. Now we assume  $1 \leq m \leq n$ . Let  $r = n \bmod m$ . Then by Lemmas 7,10, we have  $\#(m, n) = (a^m!)^{a^{n-m}-a^r} \#(m, m+r) = (a^m!)^{a^{n-m}-a^r} (a^m!)^{a^r} = (a^m!)^{a^{n-m}}$ .  $\square$

## 4 Generating Multi-Shift de Bruijn Sequences

In this section, we study the problem of generating one  $m$ -shift de Bruijn sequence of order  $n$  for arbitrary alphabet and positive integers  $m, n$ . When  $1 \leq n \leq m$ , a  $m$ -shift de Bruijn sequence of order  $n$  is easy to construct as given in Theorem 4. Now we consider the case  $1 \leq m < n$ . We will present two algorithms for generating a  $m$ -shift de Bruijn sequence of order  $n$ .

We claim that  $m$ -shift de Bruijn sequences of order  $km$  can be generated using the ordinary de Bruijn sequence generating algorithm, such as described by Fredricksen [4]. To do this, we first generate a de Bruijn sequence  $w$  of order  $k$  over the alphabet  $\Gamma = \Sigma^m$ . Then we replace each letter of  $w$  in  $\Gamma$  by the corresponding word of length  $m$  over  $\Sigma$ . It is easy to see that the new word is a  $m$ -shift de Bruijn sequence of order  $km$ .

The first algorithm of generating multi-shift de Bruijn sequence is to generate  $m_i$ -shift de Bruijn sequences of order  $k_i m_i$  for  $i = 1, 2$  before rearranging the words to obtain an arbitrary  $m$ -shift de Bruijn sequence of order  $n$ . Let  $1 \leq m < n$  be two integers, and  $n = km + r$ , where  $r = n \bmod m$ . The case  $r = 0$  is already discussed and the case  $|\Sigma| = 1$  is trivial. So we assume  $r \neq 0$  and  $|\Sigma| \geq 2$ . We define  $m_1 = r$ ,  $n_1 = (k+1)r$  and generate  $w_1 = \tau(m_1, n_1)0^{m_1}$  such that  $\tau(m_1, n_1)$  is a  $m_1$ -shift de Bruijn sequence of order  $n_1$  and  $w_1[1..n_1] = 0^{n_1}$ ; and define  $m_2 = m - r$ ,  $n_2 = k(m - r)$  and generate  $w_2 = \tau(m_2, n_2)0^{m_2}$  such that  $\tau(m_2, n_2)$  is a  $m_2$ -shift de Bruijn sequence of order  $n_2$  and  $w_2[1..n_2] = 0^{n_2}$ . Let  $a = |\Sigma|$ ,  $N_1 = a^{n_1}$ ,  $N_2 = a^{n_2}$ . We define  $u_i = w_1[n_1 + (i-1)m_1 + 1..n_1 + im_1]$ ,  $u'_i = u_{1+(i \bmod (N_1-1))}$ ,  $v_i = w_2[n_2 + (i-1)m_2 + 1..n_2 + im_2]$ ,  $v'_i = v_{1+(i-1 \bmod N_2)}$ . Then the following word

$$0^n v_1 0^{m_1} v_2 \cdots v_{N_2-1} 0^{m_1} v_{N_2} u'_{(N_1-1)N_2} v'_1 u'_1 v'_2 u'_2 \cdots v'_{(N_1-1)N_2-1} u'_{(N_1-1)N_2-1}$$

is one  $m$ -shift de Bruijn sequence of order  $n$ , where  $v_{N_2} = 0^{km}$  and  $u'_{(N_1-1)N_2} = u_1$ .

To show the correctness, we claim that every word in  $L_1 = (0^{m_1} \Sigma^{m_2})^k 0^{m_1}$  appears in

$$w' = 0^n v_1 0^{m_1} v_2 \cdots v_{N_2-1} 0^{m_1}$$

as a factor at a modulo  $m$  position exactly once. Furthermore, since  $\gcd(N_1 - 1, N_2) = 1$ , every word in  $L_2 = (\Sigma^{m_1} \Sigma^{m_2})^k \Sigma^{m_1} \setminus L_1$  appears in

$$w'' = 0^{km} u_1 v'_1 u'_1 v'_2 u'_2 \cdots v'_{(N_1-1)N_2-1} u'_{(N_1-1)N_2-1}$$

as a factor at a modulo  $m$  position exactly once. Therefore, the generated word is indeed a  $m$ -shift de Bruijn sequence of order  $n$ .

Now, we will see an example. Consider generating a 2-shift de Bruijn sequence of order 5. Then  $m_1 = 1, n_1 = 3, m_2 = 1, n_2 = 2$  and we can obtain two words  $w_1 = 00011101000$ , which is  $\tau(1, 3)0$ , and  $w_2 = 001100$ , which is  $\tau(1, 2)0$ . So one 2-shift de Bruijn sequence of order 5 is as follows

$$\begin{aligned} & 000001_2 01_2 00_2 00_2 \ 1_1 1_2 1_1 1_2 1_1 0_2 0_1 0_2 1_1 1_2 0_1 1_2 0_1 0_2 1_1 0_2 1_1 1_2 0_1 0_2 1_1 0_2 0_1 1_2 0_1 1_2 \\ & \quad 1_1 0_2 1_1 0_2 1_1 1_2 0_1 1_2 1_1 0_2 0_1 0_2 0_1 1_2 1_1 1_2 1_1 0_2 1_1 0_2 0_1 1_2 1_1 1_2 0_1 0_2 0_1, \end{aligned}$$

where the subscripts 1 and 2 denote whether the letter is from the word  $w_1$  (words  $u_i, u'_i$ ) or from the word  $w_2$  (words  $v_i, v'_i$ ).

Now we present the second algorithm, which uses the same idea of “prefer one” algorithm [11] for generating ordinary de Bruijn sequences. Let  $m, n$  be two positive integers. The following algorithm generates a  $m$ -shift de Bruijn sequence of order  $n$ :

1. Start the sequence  $w$  with  $n$  zeros;
2. Append to the end of current sequence  $w$  the lexicographically largest word of length  $m$  such that the suffix of length  $n$  of new sequence has not yet appeared as factor at a modulo  $m$  position;
3. Repeat the last step until no word can be added.

To show the correctness, first we claim that when the algorithm stops, the suffix  $u$  of length  $n - m$  of  $w$  contains only zeros. To see this, suppose  $u$  is not  $0^{n-m}$ . Since no word can be added, all  $|\Sigma|^m$  words of length  $n$  with prefix  $u$  appear in  $w$  and thus  $u$  appears in  $w$  as a factor at a modulo  $m$  position  $|\Sigma|^m + 1$  times. So there are  $|\Sigma|^m + 1$  words of length  $n$  with suffix  $u$  that appear in  $w$  at a modulo  $m$  position, which contradicts the definition of the multi-shift de Bruijn sequence. Therefore,  $u = 0^{n-m}$ . Furthermore, word  $0^{n-m}$  appears in  $w$  as a factor at a modulo  $m$  position  $|\Sigma|^m + 1$  times and thus all words in  $\Sigma^m 0^{n-m}$  appear in  $w$  as a factor at a modulo  $m$  position. By the algorithm, no word of length  $n$  can appear twice in  $w$  at a modulo position. So, in order to prove the correctness of the algorithm, it remains to show every word of length  $n$  appears in  $w$  as a factor at a modulo  $m$  position. Suppose a word  $v$  does not appear in  $w$  at a modulo  $m$  position. Then  $v[m + 1 .. n] \neq 0^{n-m}$  and the word  $v[m + 1 .. n]0^m$  does not appear in  $w$  as a factor at a modulo  $m$  position as well; otherwise, there are  $|\Sigma|^m$  appearance of  $v[m + 1 .. n]$  in  $w$  at a modulo  $m$  position, which means  $v$  appears in  $w$  as a factor at a modulo  $m$  position. Repeat this procedure, none of the words  $v[m + 1 .. n]0^m, v[2m + 1 .. n]0^{2m}, \dots, v[\lfloor n/m \rfloor m + 1 .. n]0^{\lfloor n/m \rfloor m}$  appears in  $w$  as a factor at a modulo  $m$  position. But for  $\lfloor n/m \rfloor m \geq n - m$ , we proved that  $v[\lfloor n/m \rfloor m + 1 .. n]0^{\lfloor n/m \rfloor m}$  appears in  $w$  as a factor at a modulo  $m$  position, a contradiction. Therefore, every word of length  $n$  appears at a modulo  $m$  position.

Now, we use the algorithm to generate one 2-shift de Bruijn sequence of order 5. Starting from 00000, since 00011 does not appear as a factor at a modulo 2 position, we append 11 to the current sequence 00000. Repeating this procedure and appending words 11, 11, 10, 11,  $\dots$ , finally we obtain the word:

0000011111110111010110111011001110011001010011000100001010100010000

If we circularly move the prefix  $0^n$  to the end, the sequence generated by the second algorithm is the lexicographically largest  $m$ -shift de Bruijn sequence of order  $n$ .

## 5 Application in the Frobenius Problem in a Free Monoid

The study of multi-shift de Bruijn sequences is inspired by a problems of words, called the Frobenius problem in a free monoid. Given  $k$  integers  $x_1, \dots, x_k$ , such that  $\gcd(x_1, \dots, x_k) = 1$ , then there are only finitely many positive integers that *cannot* be written as a non-negative integer linear combination of  $x_1, \dots, x_k$ . The integer *Frobenius problem* is to find the largest such integer, which is denoted by  $g(x_1, \dots, x_k)$ . For example,  $g(3, 5) = 7$ .

If words  $x_1, \dots, x_k$ , instead of integers, are given such that there are only finitely many words that *cannot* be written as concatenation of words from the set  $\{x_1, \dots, x_k\}$ , the *Frobenius problem in a free monoid* [8] is to find the longest such words. If all  $x_1, \dots, x_k$  are of length either  $m$  or  $n$ ,  $0 < m < n$ , there is an upper bound: the length of the longest word that cannot be written as concatenation of words from the set  $\{x_1, \dots, x_k\}$  is less than or equal to  $g(m, l) = ml - m - l$ ,

where  $l = m\Sigma^{n-m} + n - m$ . [8] Furthermore, the upper bound is tight and the construction is based on the multi-shift de Bruijn sequences. We denote the set of all words that can be written as concatenation of words in  $S$ , including the empty word, by  $S^*$ .

**Theorem 12.** [8] *There exists  $S \subseteq \Sigma^m \cup \Sigma^n$ ,  $0 < m < n$ , such that  $\Sigma^* \setminus S^*$  is finite and the longest words in  $\Sigma^* \setminus S^*$  constitute exactly the language  $(\tau\Sigma^m)^{m-2}\tau$ , where  $\tau$  is a  $m$ -shift de Bruijn sequence of order  $n - m$ .*

For example, for any set of words  $S \subseteq U = \{0, 1\}^3 \cup \{0, 1\}^7$  such that  $\{0, 1\}^* \setminus S^*$  is finite, the longest words in  $\{0, 1\}^* \setminus S^*$  are of length less than or equal to  $g(3, 3 \cdot 2^4 + 4) = g(3, 52) = 101$ . To construct  $S$  to reach the upper bound, we first choose an arbitrary 3-shift de Bruijn sequence of order 4 as  $\tau = 000011111110110101101100100011011010010001001000$ . Then based on  $\tau$ , we construct the set  $S = U \setminus \{0000111, 0111111, 1111110, 1110110, 0110101, 0101101, 1101100, 1100100, 0100011, 0011011, 1011010, 1010010, 0010001, 0001001, 1001000\}$ . We have  $L = \{0, 1\}^* \setminus S^* = \tau\{0, 1\}^2\tau$  and one of the longest words in  $L$  of length exactly 101 is given below:

0000111111101101011011001000110110100100010010001  
11000011111110110101101100100011011010010001001000.

## 6 Conclusion

In this paper, we generalized the classic de Bruijn sequence to a new multi-shift setting. A word  $w$  is a  $m$ -shift de Bruijn sequence  $\tau(m, n)$  of order  $n$ , if each word of length  $n$  appears exactly once as a factor at a modulo  $m$  position. An ordinary de Bruijn sequence is a 1-shift de Bruijn sequence.

We showed the total number of distinct  $m$ -shift de Bruijn sequences of order  $n$  is  $\#(m, n) = (a^n)!a^{(m-n)(a^n-1)}$  for  $1 \leq n \leq m$  and is  $\#(m, n) = (a^m!)^{a^{n-m}}$  for  $1 \leq m \leq n$ , where  $a = |\Sigma|$ . This result generalizes the formula  $(a!)^{a^{n-1}}$  for the number of ordinary de Bruijn sequences [1]. Here we use an ordinary word form; if counting the sequences in a circular form, then the number is to be divided by  $a^n$ .

We provided two algorithms for generating a  $m$ -shift de Bruijn sequence of order  $n$ . The first algorithm is to rearrange factors from two simpler multi-shift de Bruijn sequences, where the order is a multiple of the shift. The second is the analogue of the “prefer one” algorithm (for example, see [4]) for generating ordinary de Bruijn sequence.

The multi-shift de Bruijn sequence has application in the Frobenius problem in a free monoid by providing constructions of examples. It will be interesting to see that this generalized concept of the de Bruijn sequence can help in other fields of theoretical computer science and discrete mathematics.

## Acknowledgements

The author would like to thank Prof. Jeffrey Shallit for valuable discussion.

## References

- [1] T. van Aardenne-Ehrenfest and N. G. de Bruijn. Circuits and trees in oriented linear graphs. *Simon Stevin*, 28:203–217, 1951.



- [2] N. G. de Bruijn. A combinatorial problem. *Indag. Math.*, 8(4):461–467, 1946.
- [3] C. Flye Sainte-Marie. Solution to question nr. 48. *L’Intermédiaire Math.*, 1:107–110, 1894.
- [4] H. Fredricksen. A survey of full length nonlinear shift register cycle algorithms. *SIAM Review*, 24(2):195–221, 1982.
- [5] H. Fredricksen and I. J. Kessler. Lexicographic compositions and de bruijn sequences. *J. Combin. Theory Ser. A*, 22:17–30, 1977.
- [6] S. W. Golomb. *Shift Register Sequences*. Holden-Day, 1967.
- [7] I. J. Good. Normal recurring decimals. *J. London Math. Soc.*, 21(3):167–169, 1946.
- [8] J.-Y. Kao, J. Shallit, and Z. Xu. The Frobenius problem in a free monoid. In *STACS 2008, Proc. 25th Internat. Symp. Theoretical Aspects of Comp. Sci.*, pages 421–432, 2008.
- [9] G. Kirchhoff. Über die Auflösung der Gleichungen, auf welche man bei der untersuchung der linearen verteilung galvanischer Ströme geführt wird. *Ann. Phys. Chem.*, 72:497–508, 1847.
- [10] D. E. Knuth. *The Art of Computer Programming*. Addison-Wesley, 1969.
- [11] M. H. Martin. A problem in arrangements. *Bull. Amer. Math. Soc.*, 40:859–864, 1934.
- [12] A. Ralston. De Bruijn sequences — a model example of the interaction of discrete mathematics and computer science. *Math. Mag.*, 55(3):131–143, 1982.
- [13] D. Rees. Note on a paper by I. J. Good. *J. London Math. Soc.*, 21(3):169–172, 1946.
- [14] W. T. Tutte and C. A. B. Smith. On unicursal paths in a network of degree 4. *Amer. Math. Monthly*, 48(4):233–237, 1941.